

National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption

Gergely Buda (BSE)

Vasco M. Carvalho (University of Cambridge & CEPR)

Stephen Hansen (Imperial College & CEPR)

Alvaro Ortiz (BBVA Research)

Tomasa Rodrigo (BBVA Research)

José V. Rodríguez Mora (University of Edinburgh & CEPR)

Conference on Non-traditional Data, Machine Learning and NLP in Macroeconomics, Co-Hosted by
Sveriges Riksbank, Federal Reserve Board, the Bank of Italy and the Bank of Canada.

Stockholm, October 3-4, 2022

Link to [CEPR paper](#)

Link to [Cambridge WP in Economics](#)

Why Use Transaction Data

- **Modern payment systems** generate a vast amount of real-time data on activity which can be **incorporated into national accounting measures** (Bean 2016).
 - Advantages in term of **Real Time, Granularity and Cost** to statistical agencies (albeit not to private sector!!).
 - The **High Definition component** open the door for the **design of Smart Policies** (addressing policy where more needed and/or higher impact)
 - Growing **interest by Statistical Agencies to incorporate “Transaction & Registry data” to improve Statistical Accounts**
 - But few, if any, attempts in the academic literature to **build national accounting objects from first principles using large-scale payment data.**
-

Contributions: Four contributions of the Paper

1. How to **construct representative panel of household expenditure**. Massive survey.
2. Show that it **aggregates to Quarterly National Accounts**
3. Create **Distributional National Accounts for Consumption**
4. Study the **Micro-Structure of consumption dynamics** (just an example !!):

First proof of concept that naturally occurring transaction data, arising through the millions of transactions of economic agents, can be organized via national accounting rules and harnessed to produce a large-scale, high-quality and highly-detailed consumption survey

Building a Massive Consumption Survey

Building a Consumption Survey: Data & Key Problems

- Universe of BBVA retail accounts in Spain by BBVA
- Allowing us to track expenditure as it flows out of these accounts, transaction by transaction
- 3 billion individual transactions by 1.8 million BBVA customers, from 2016 to 2021

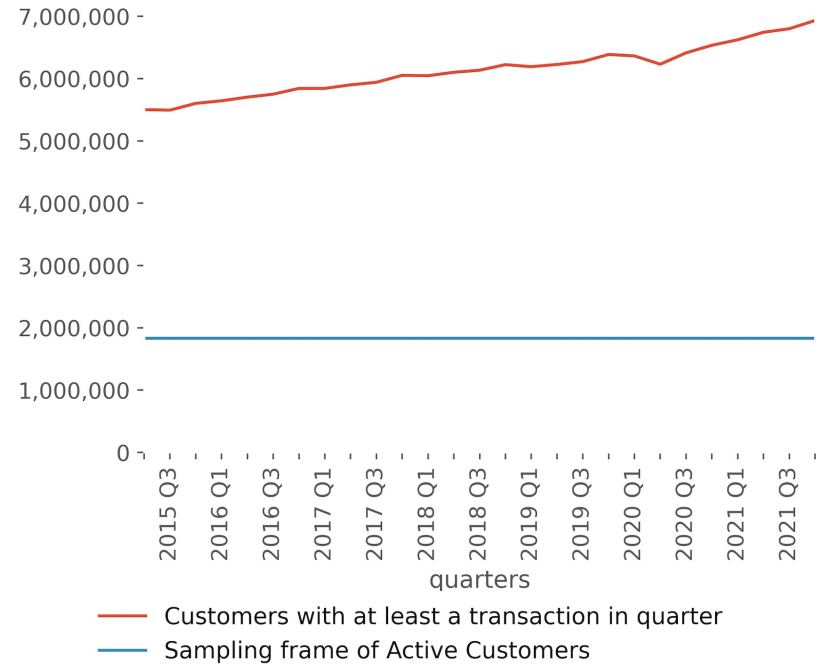
- Two problems to translate transactions Bank data into a population sample:
 - The client pool of a **Bank** \neq **Population: Biases**
 - Spending is not the same than consumption: **Spending** \neq **Consumption**
-

Building a Consumption Survey: Active Customers

Sample Frame

- 10,270,041 unique customers (2015-2021)
- Most spend infrequently or for short periods only.
- Define "Active Customers" as making at least 10 consumption related transactions in each quarter.
- 1,827,866

Customers with transactions and sampling frame of active customers



Building a Consumption Survey: Non-Housing Expenditures

Classification of Non-Housing Consumption Spending in COICOPS

- If a transaction is explicitly categorized in one of the 12 COICOPS.
- Follow national accounting principles wherever possible

Card Data

- Merchant Client Code (MCC) of the counterparty firm.
- Manual Mapping to COICOPS
- Multi-product retailers. Assigned by external data on distributions.

Transfers

- String match counter-party name to commercial registry.
- If counter-party is located as a firm, we assign as above.

Direct Debit

- ~ 100 internal labels.
- Manual Mapping
- When this is unclear, we read field, determine firm and use either MCC (if possible) or NACE code of firm to assign COICOP.

Cash

- Both cash and over the counter.
- Assume is consumption.
- Assumptions on distribution.

Spending Category	Volume of Transactions	Number of Transactions
Offline Card Transactions	60,319 million	1,772 million
Online Card Transactions	11,858 million	313 million
Direct Debits	66,036 million	752 million
Cash Withdrawal	64,592 million	359 million
Transfers excl. rent	11,148 million	15 million

Building a Consumption Survey: Housing Consumption (Rents)

Determination of Housing Consumption Spending

- We locate payment of rental for housing services.
 - Reading of free-text field in direct debits and transfers.
 - Minimum 100 EUR
 - Exclude parking, etc.
 - Payments made in 70 months.
 - 32,127 households.
- Use household covariates to predict monthly rent
 - Income (from BBVA table, six month average)
 - Utility Payments (direct debits)
 - Geography: 327 regions (consolidating postal codes)

Variable	Model	Test set
Spending on House Utilities	0.0884 (0.0008)	
Income	0.0362 (0.0011)	
N of Contract Groups	16,977	15,512
N of Observations	1,134,735	15,512
R ²	0.3911	
Adjusted R ²	0.3765	
Within R ²	0.1200	
Root MSE	204.6144	221.64

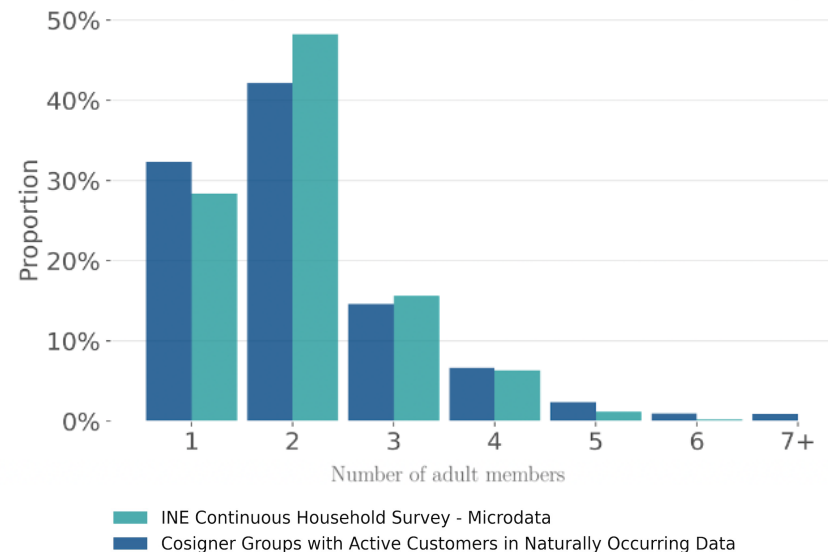
Out-of-sample behavior is reasonable with households that are 50-70 months in data.

Use covariates to IMPUTE housing consumption for the rest of the households (the vast majority)

Building a Consumption Survey: Estimating Households

- Link clients into perceived household groups.
- Individuals with whom they share a contract and live in same postal code.
- 1,589,280 household groups

Households Proxy vs Official Data
(Customers with transactions and sampling frame of active customers)



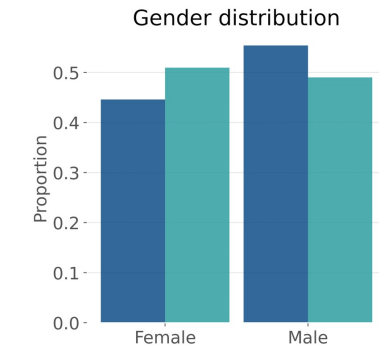
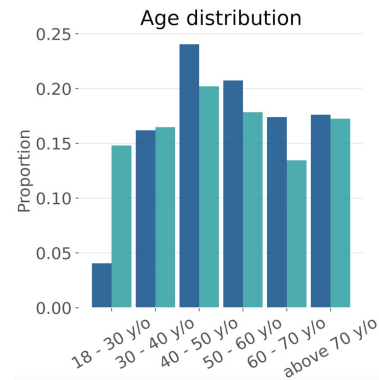
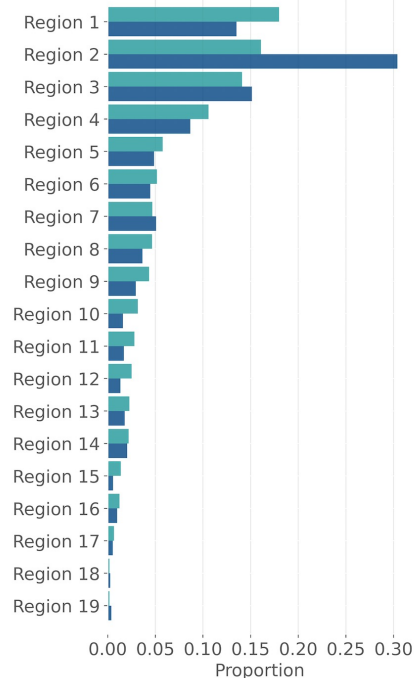
	HBS 2016	HBS 2017	HBS 2018	HBS 2019	HBS 2020	BBVA Sample
Households	22,011	22,043	21,395	20,817	19,170	1,589,280
Adults	47,420	47,055	45,328	43,988	40,285	1,827,866

Building a Consumption Survey: Demographics of Active Clients

Demographics of Active Customers

$$c_{g,a,r} = c_{g,a,r}^{hh} \left(\frac{x_{g,a,r}^{INE}}{x_{g,a,r}^{BBVA}} \right)$$

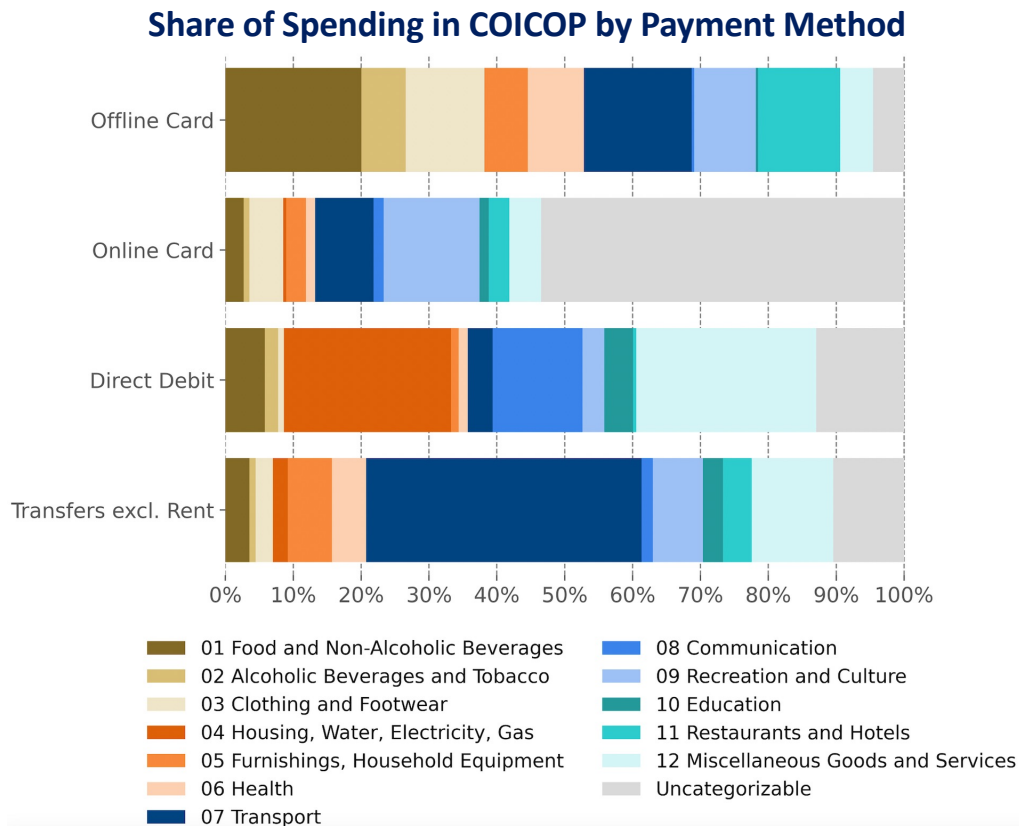
Regional distribution (Autonomous Communities)



Building a Consumption Survey: Payment-Consumption Connection

Connecting Payments with Consumption

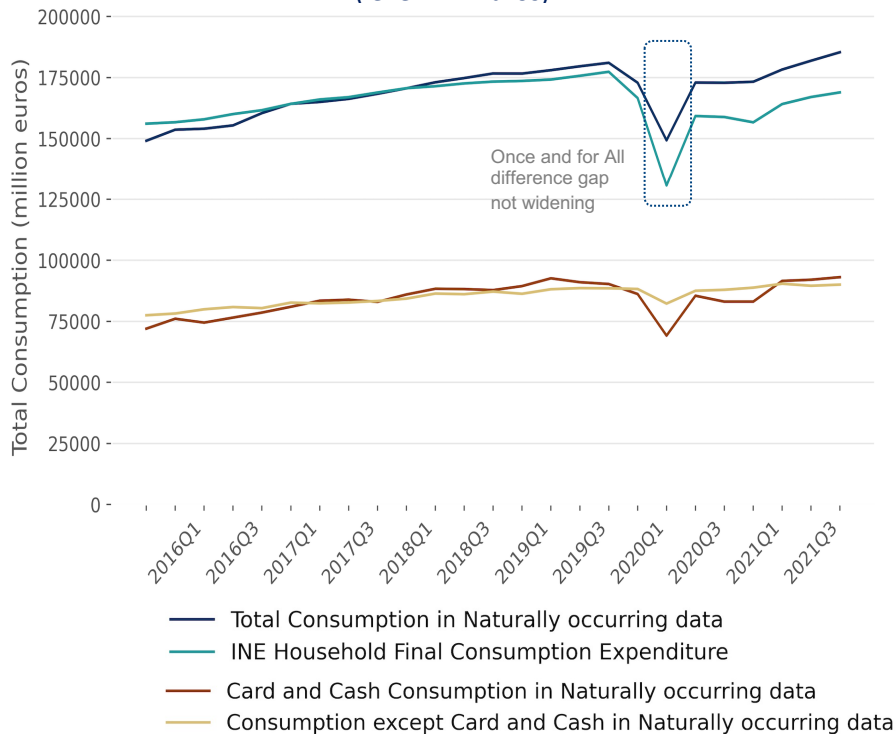
COICOP Shares by Payment Method



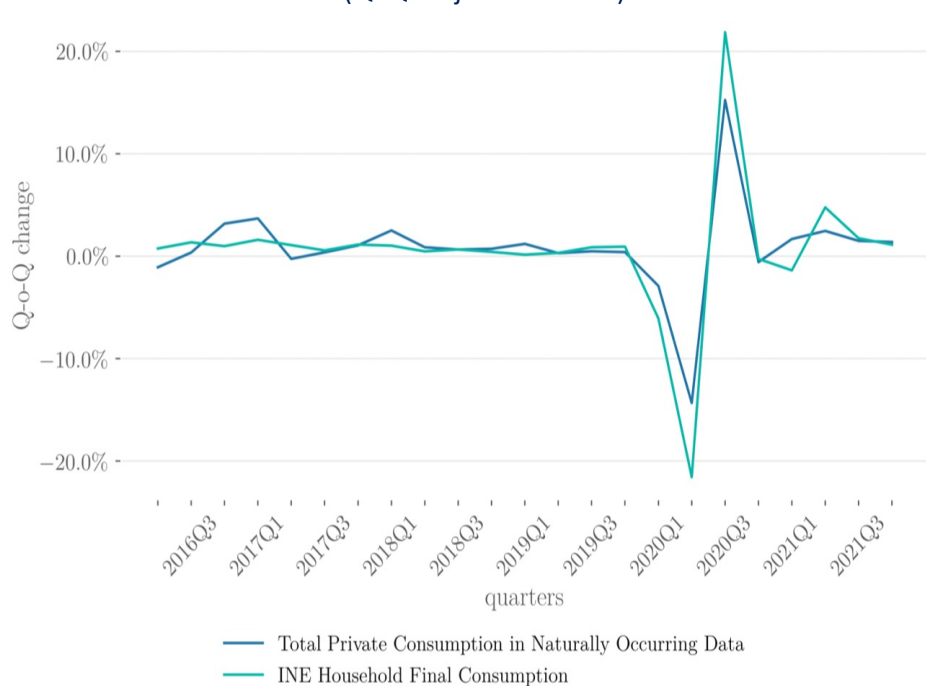
Aggregating To National Accounts

Remarkable Similar Levels & Growth from vastly different methods

Total Consumption: Naturally Occurring vs Official (level Mill Euros)



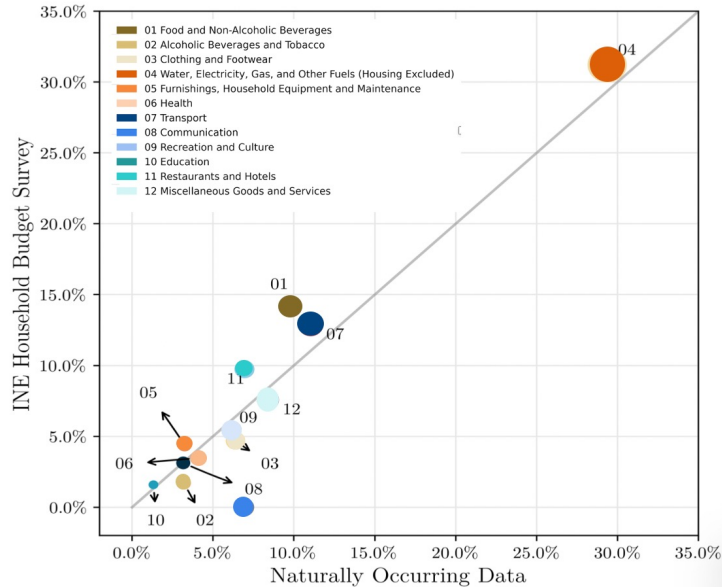
Total Consumption: Naturally Occurring vs Official (QoQ Sadj Growth Rate)



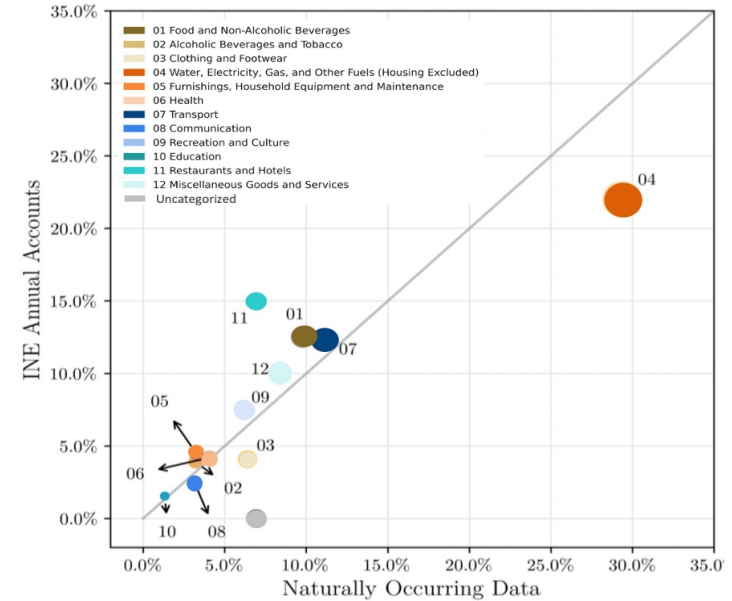
National Accounts: COICOP Distributions in HBS and N. Accounts

Distribution of Spending across COICOP Categories

(a) BBVA vs HBS

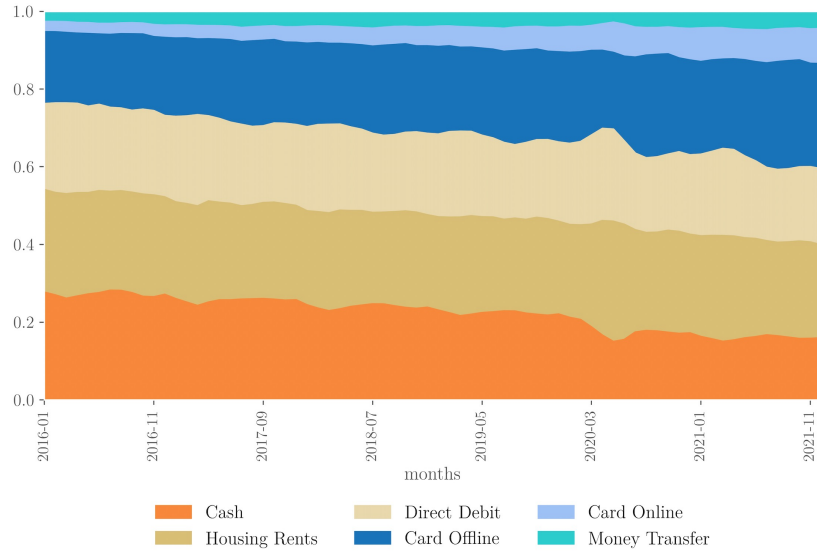


(b) BBVA vs National Accounts

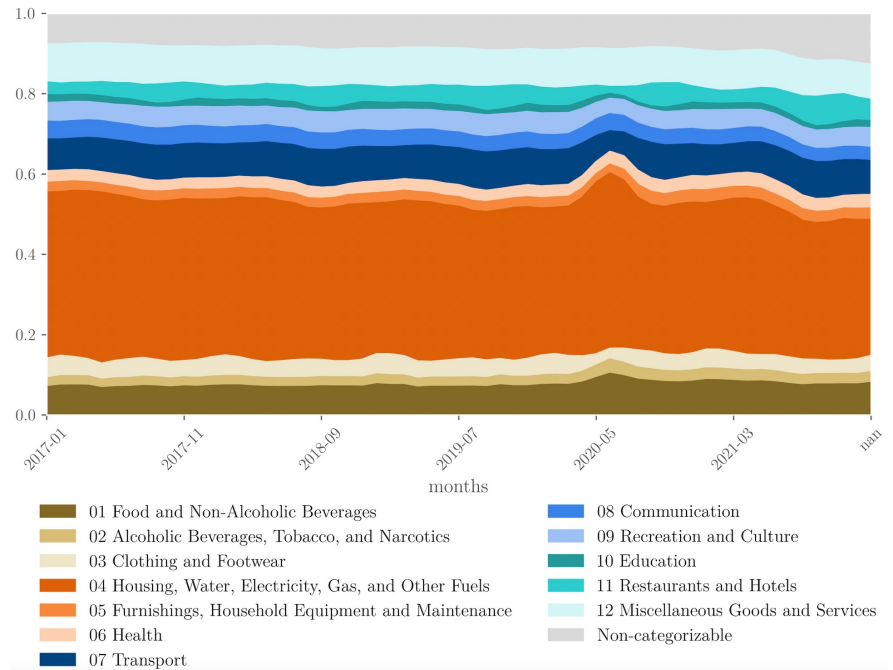


National Accounts: Richer Dynamics (Daily) by Payment & Category

Proportions of Consumption by Payment Method (Cumulative 3M, % of Total)



Proportions of Consumption by COICOP (Cumulative 3M, % of Total)

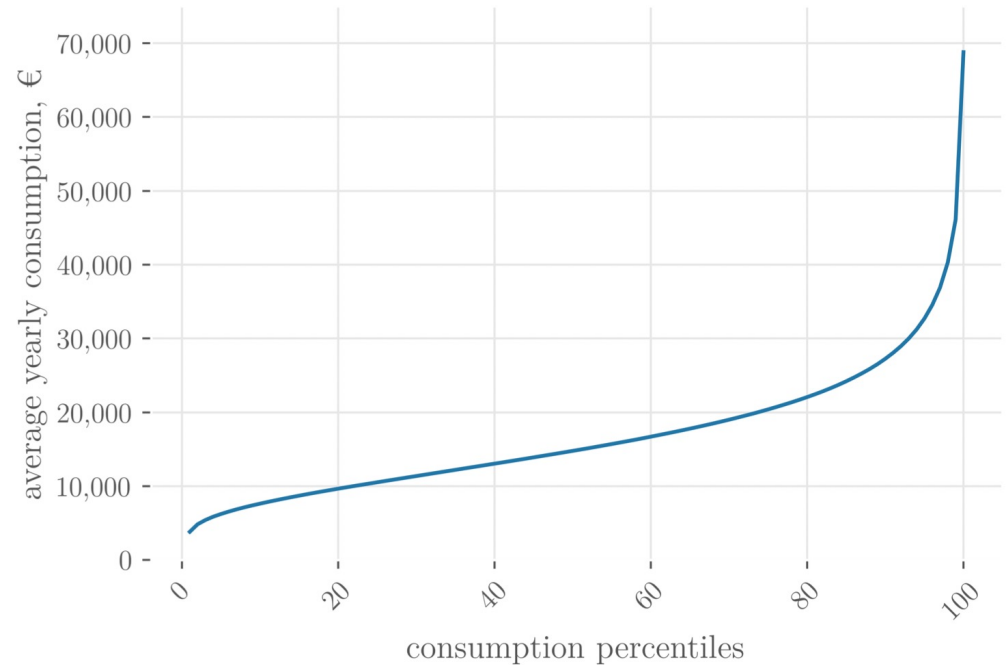


Creating Distributional National Accounts For Consumption

Distribut. Accounts: Consumption's Macro-Consistent distribution

- Macro-consistent, distribution of Consumption.
- It aggregates into NA
- Follows Picketty et al 2018.

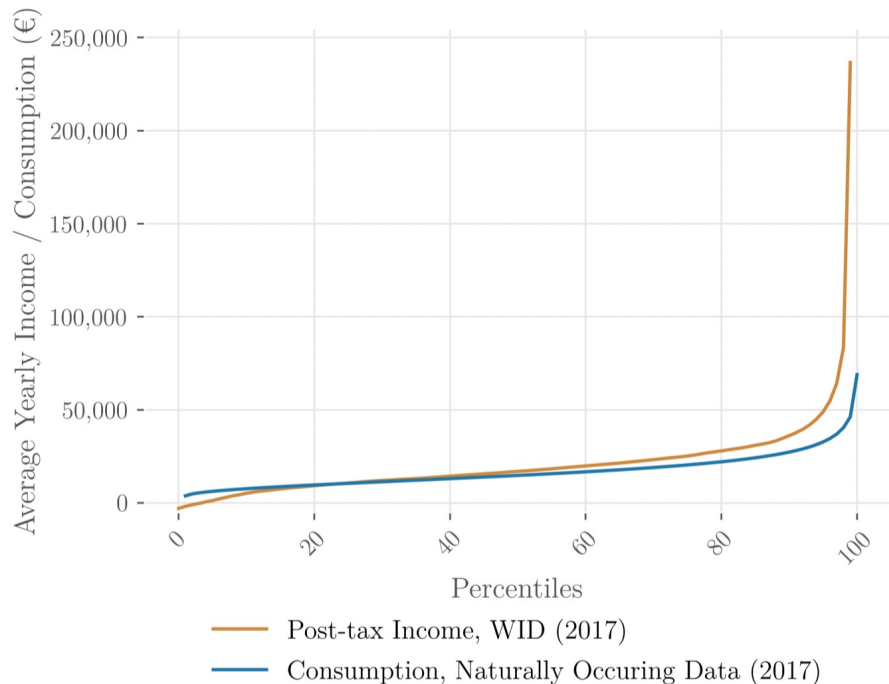
Yearly adult consumption by levels of consumption in 2017;
(Weighted sampling procedure, 2017 Euro by percentiles of consumption)



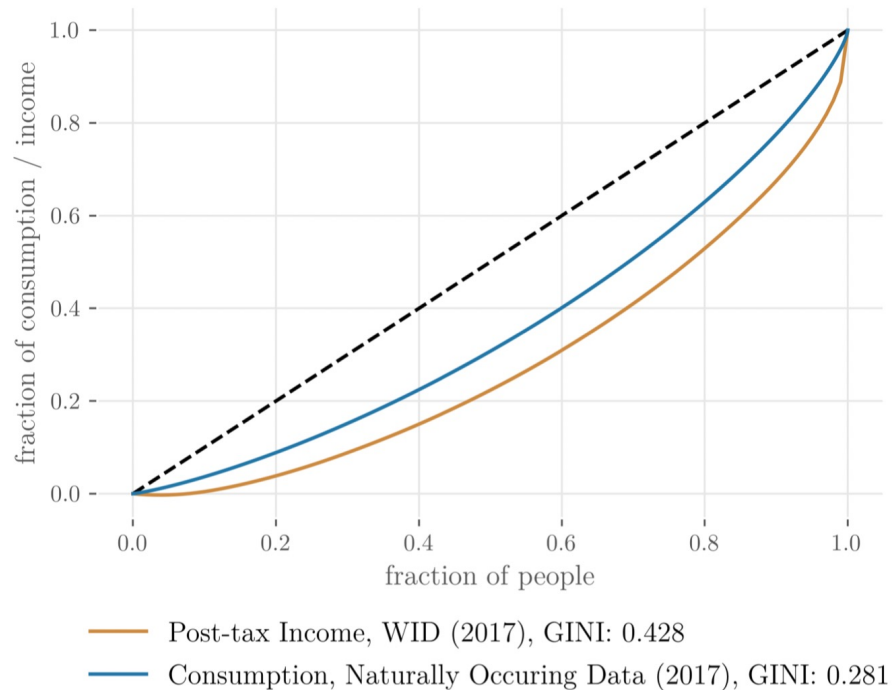
Source: Buda et Al (2022).

Distributional Accounts: Differences Consumption and Income

**Consumption Distribution per Spanish adult in 2017
BBVA vs. WID 2017 post-tax income distribution.**



**Lorenz Curves of BBVA 2017 consumption
and WID 2017 post-tax income in Spain.**



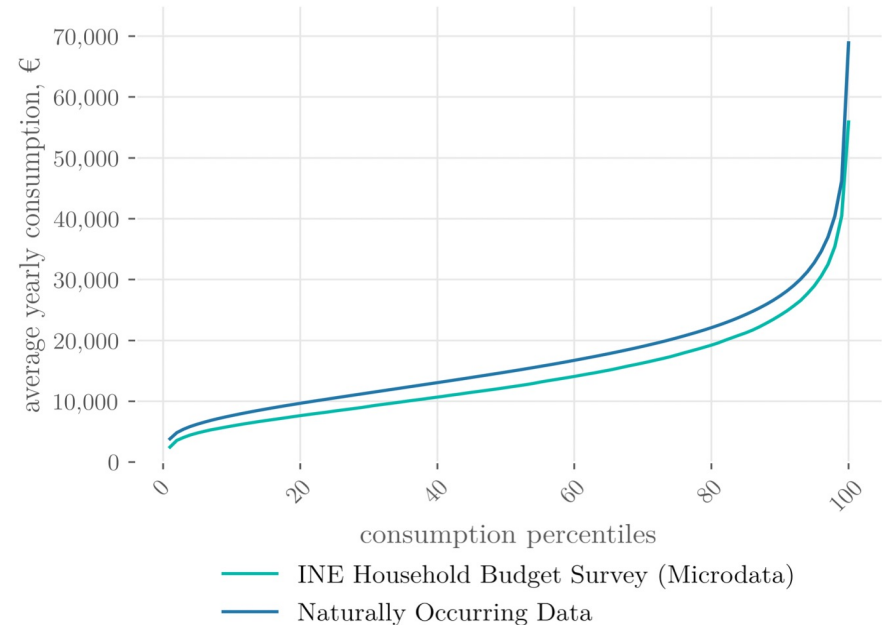
Distributional Accounts: Right Tail Difference (in Level)

Comparison Distribution of Consumption with Household Business Survey (HBS)

- Remember: difference in level.
- Right tail difference: Bigger share of consumption among the people who consumes most.

Distribution of Consumption BBVA vs. Spanish Household Budget Survey

(yearly consumption per adult distribution 2017)



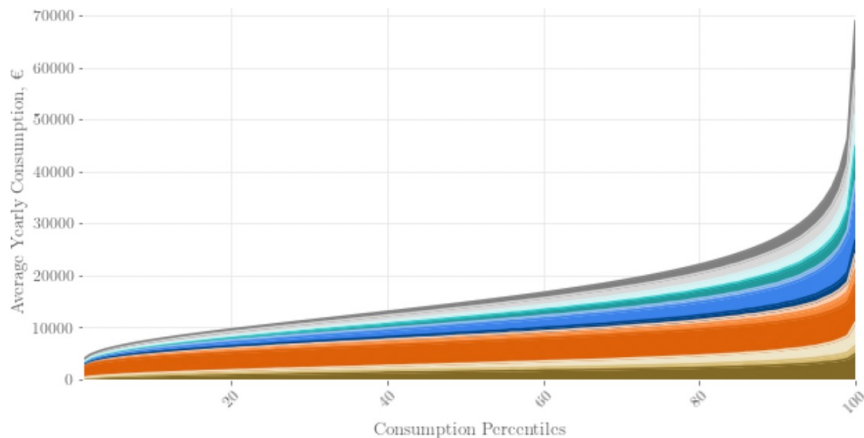
Source: Buda et Al (2022).

Distributional Accounts: Different Categories by Consumers

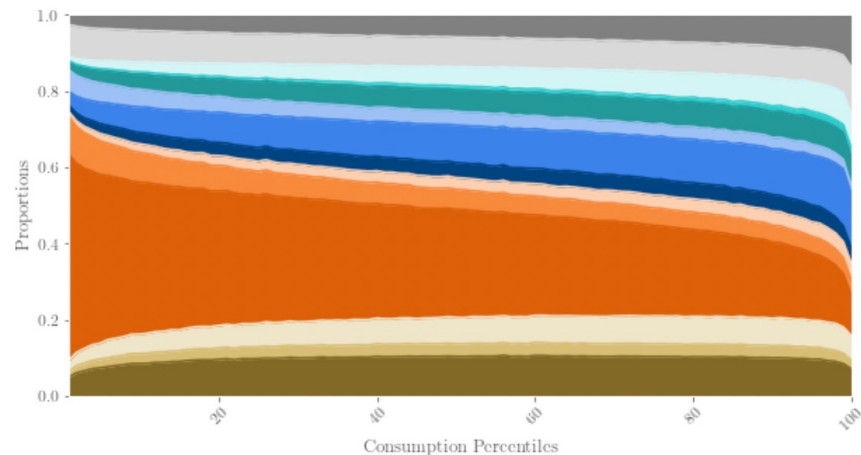
Distribution across COICOPs and People

Spain: Consumption distribution disaggregated by COICOP consumption categories

Levels



Share

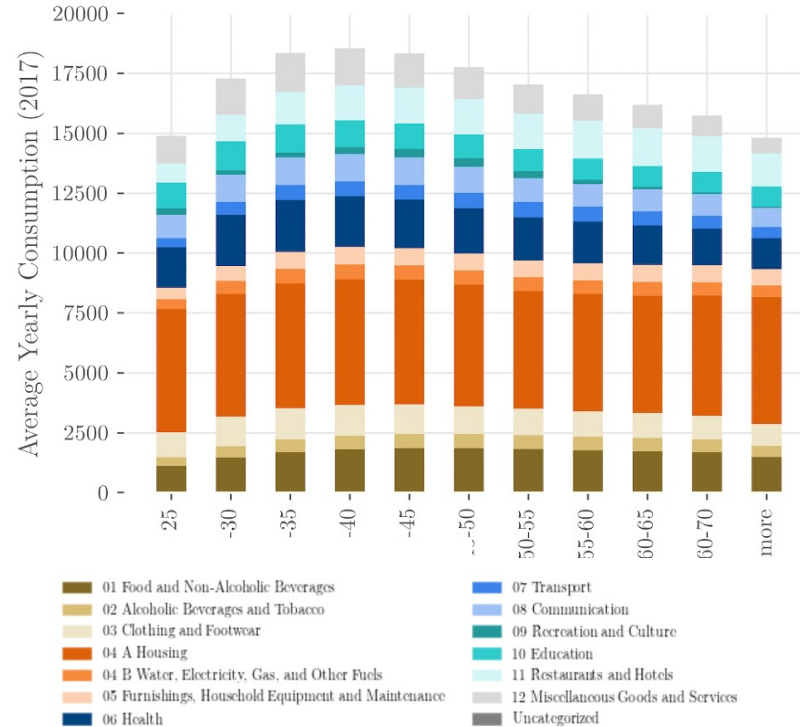


- 01 Food and Non-Alcoholic Beverages
- 02 Alcoholic Beverages and Tobacco
- 03 Clothing and Footwear
- 04 A Housing
- 04 B Water, Electricity, Gas, and Other Fuels
- 05 Furnishings, Household Equipment and Maintenance
- 06 Health
- 07 Transport
- 08 Communication
- 09 Recreation and Culture
- 10 Education
- 11 Restaurants and Hotels
- 12 Miscellaneous Goods and Services
- Uncategorized

- 01 Food and Non-Alcoholic Beverages
- 02 Alcoholic Beverages and Tobacco
- 03 Clothing and Footwear
- 04 A Housing
- 04 B Water, Electricity, Gas, and Other Fuels
- 05 Furnishings, Household Equipment and Maintenance
- 06 Health
- 07 Transport
- 08 Communication
- 09 Recreation and Culture
- 10 Education
- 11 Restaurants and Hotels
- 12 Miscellaneous Goods and Services
- Uncategorized

Distributional Accounts: Different Categories by Consumers

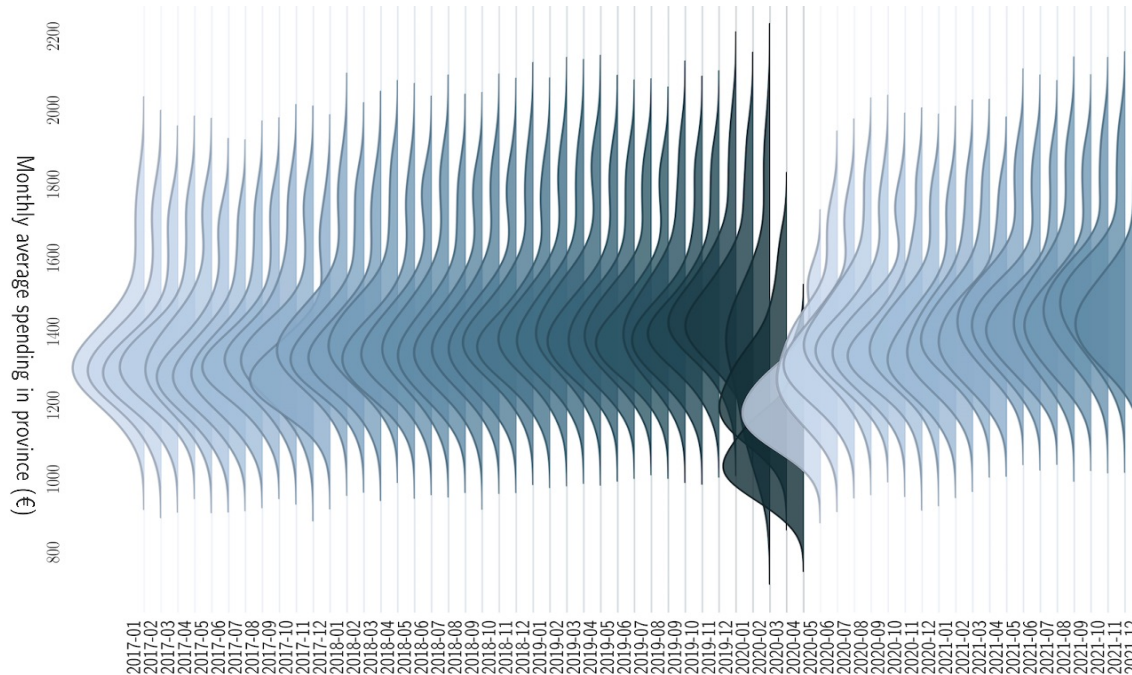
Spain: Consumption distribution by COICOP consumption categories & Age



Source: Buda et Al (2022).

Distributional Accounts: Spending across Provinces over time

Spain: Distribution of Household Consumption trough Provinces (2017-2021)
(Distribution of Average Monthly Spending by Euros)

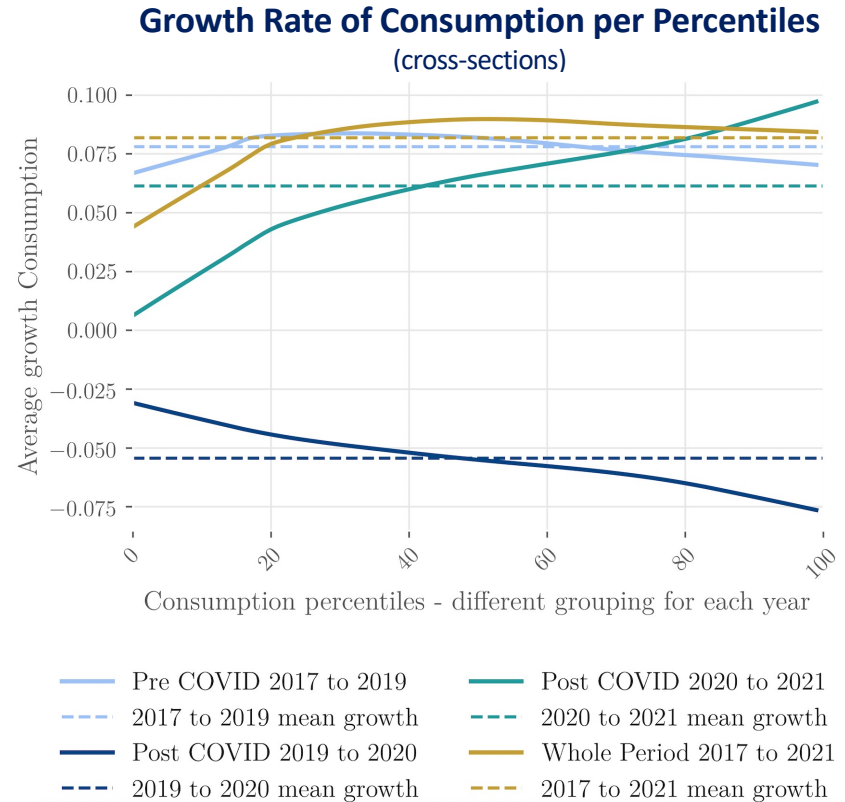


Source: Buda et Al (2022).

Distributional Accounts: Consumption Growth Distribution

Distribution of Consumption Growth

- Consumption Growth Inequality not so evident in normal years (2017-2019)
- But increasingly apparent during Covid (2019-2020)...
- ..and Post-Covid Recovery (2020-2021)



Source: Buda et Al (2022).

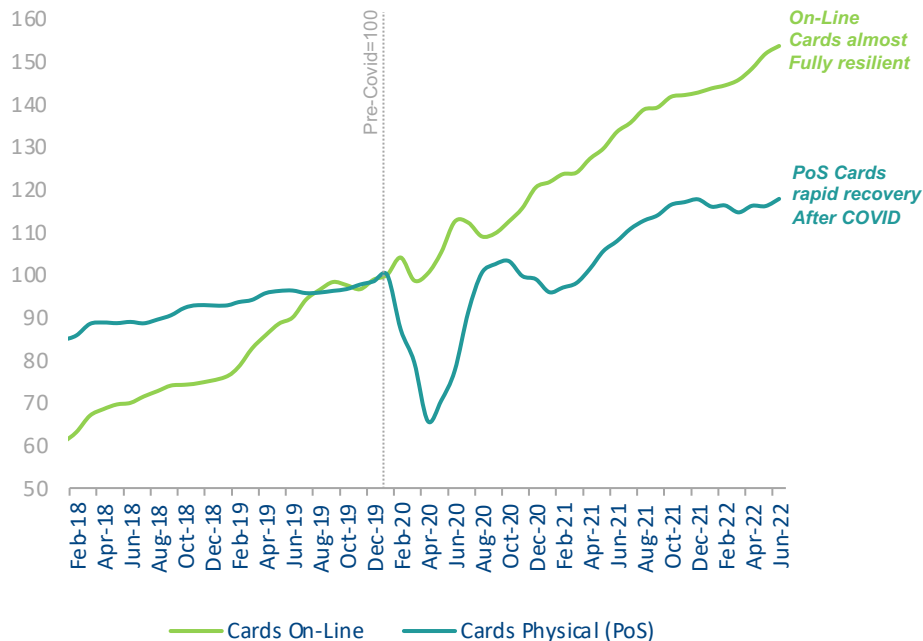
Potential Uses:

The Economy in Real
Time, High Definition and
Design of Smart Policies

Cards payments (PoS & OnLine) proved to be very useful to track activity

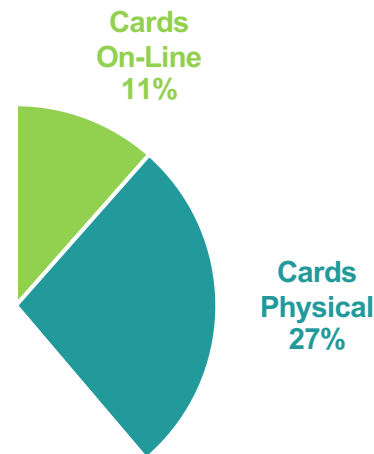
Spain: Evolution of Payments for Consumption

(SAdj. Cum 3M January 2020=100)



Spain: Distribution of Payments for Consumption

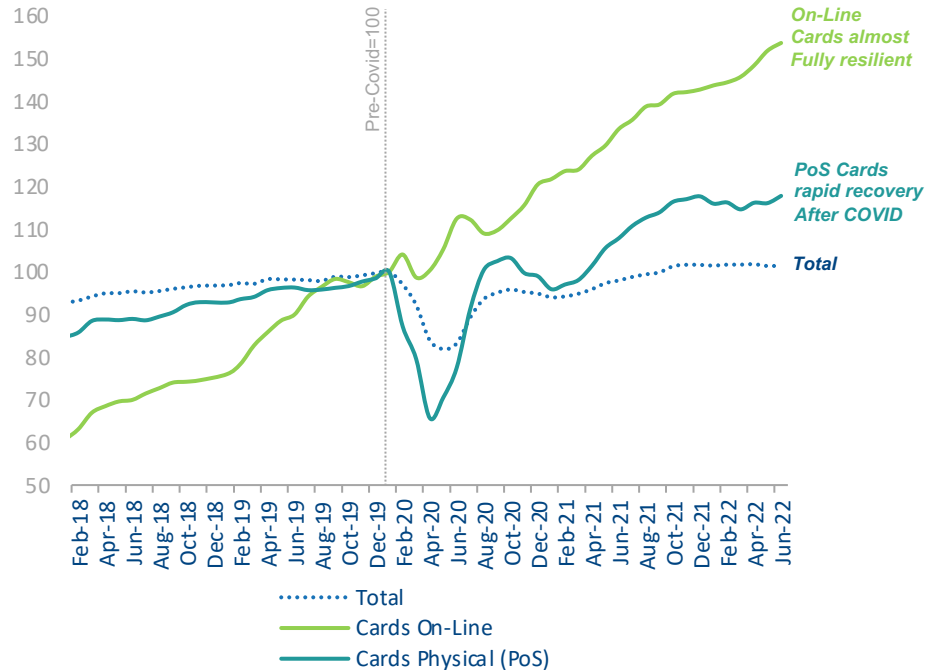
(Cumulative June 2020-2021, % of Total)



But this view is partial (38%) and biased (upwards) after the COVID..

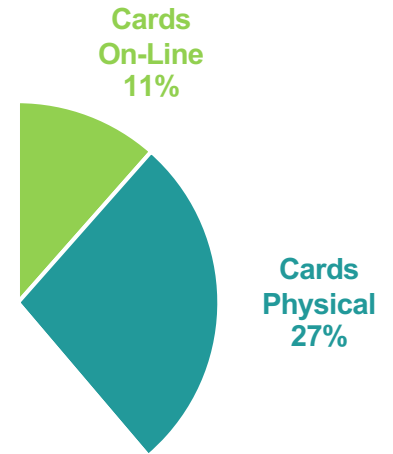
Spain: Evolution of Payments for Consumption

(SAdj. Cum 3M January 2020=100)



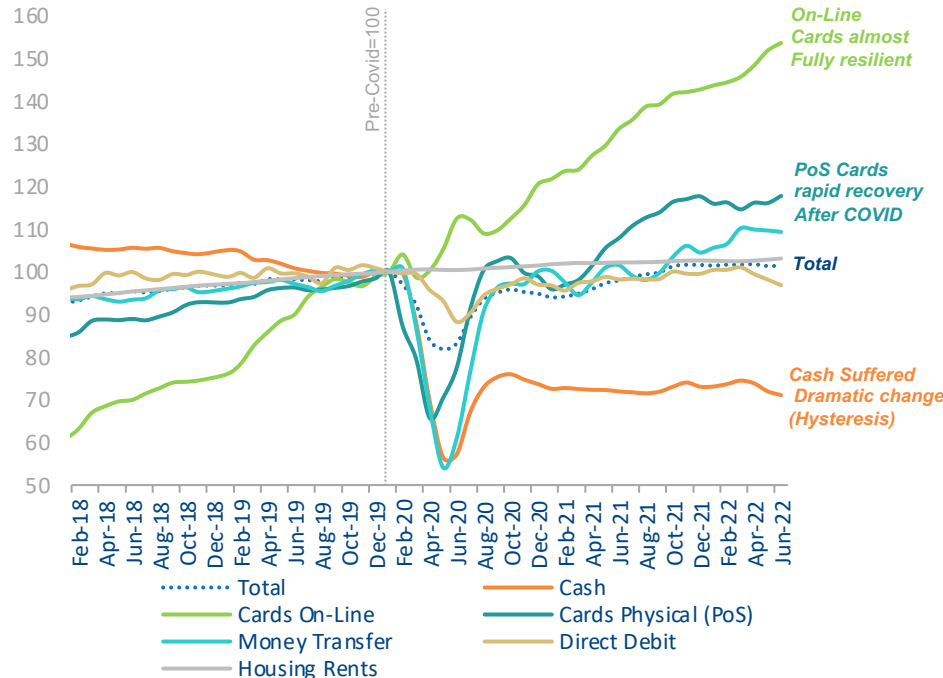
Spain: Distribution of Payments for Consumption

(Cumulative June 2020-2021, % of Total)

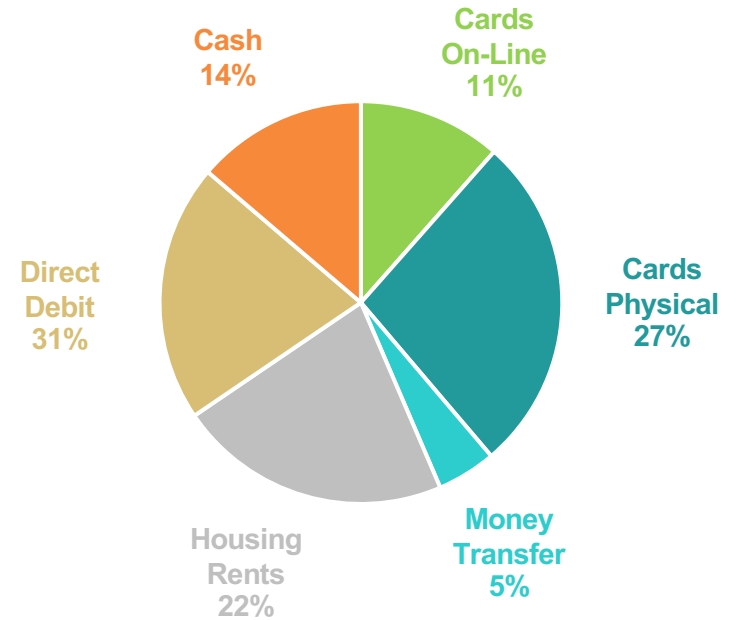


Cards can be misleading... Leading to an optimistic view

Spain: Evolution of Payments for Consumption
(SAdj. Cum 3M January 2020=100)



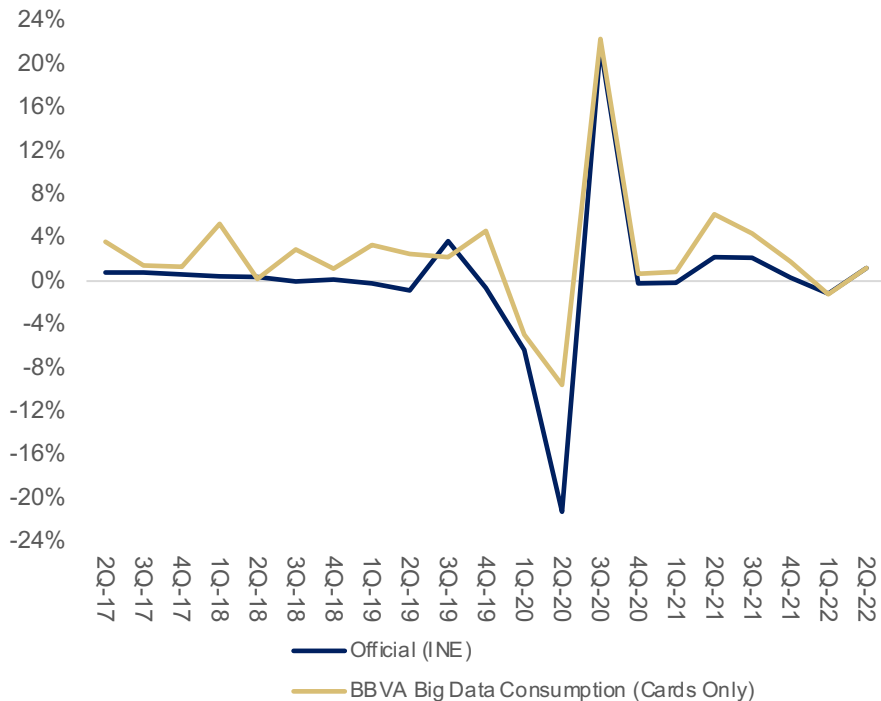
Spain: Distribution of Payments for Consumption
(Cumulative June 2020-2021, % of Total)



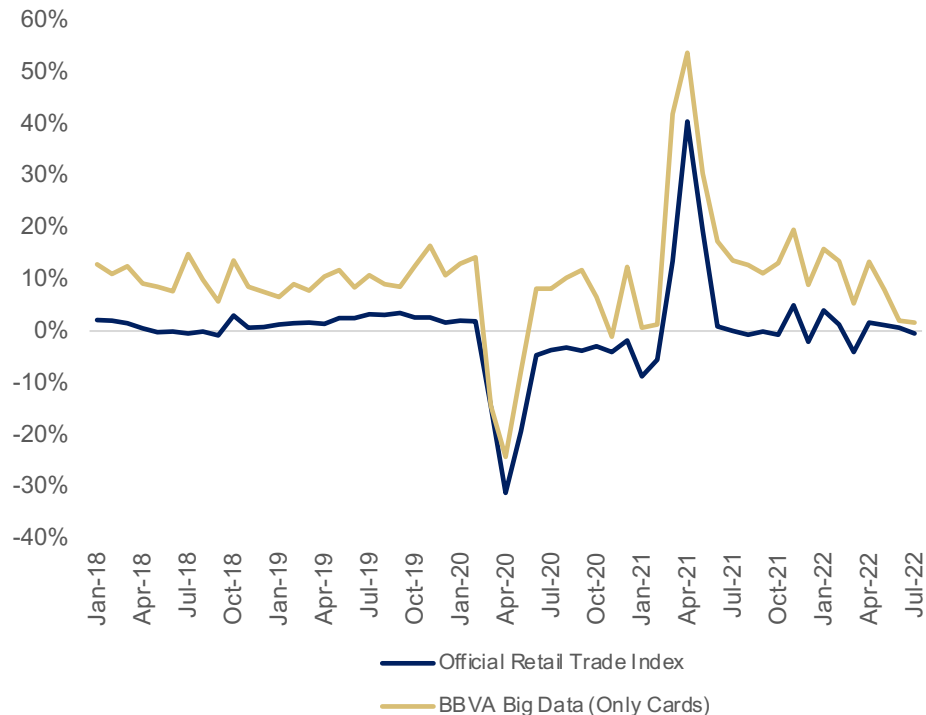
Source: BBVA Research & Buda et Al (2022).

How Naturally Occurring fits Official Statistics: The case of Cards

Spain: Real Household Consumption vs BBVA BigData
(Sadj %QoQ Growth Rate CPI deflated)

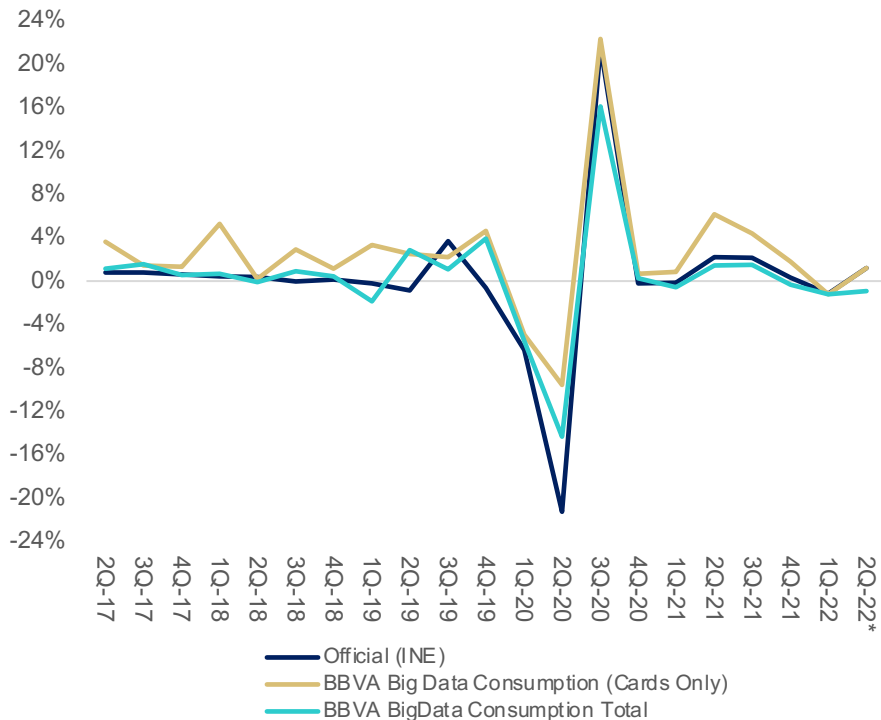


Spain: Retail Sales Index vs BBVA Big Data
(Sadj %mom Growth Rate. Big Data CPI deflated)

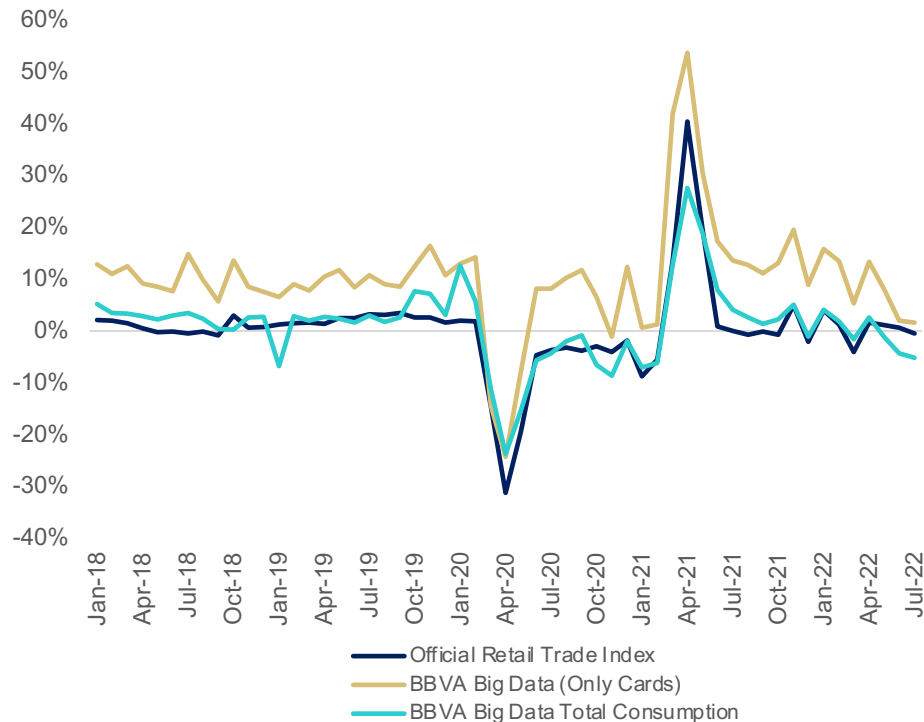


How Naturally Occurring fits Official Statistics: Total Consumption

Spain: Real Household Consumption vs BBVA BigData
(Sadj %QoQ Growth Rate CPI deflated)



Spain: Retail Sales Index vs BBVA Big Data
(Sadj %mom Growth Rate. Big Data CPI deflated)



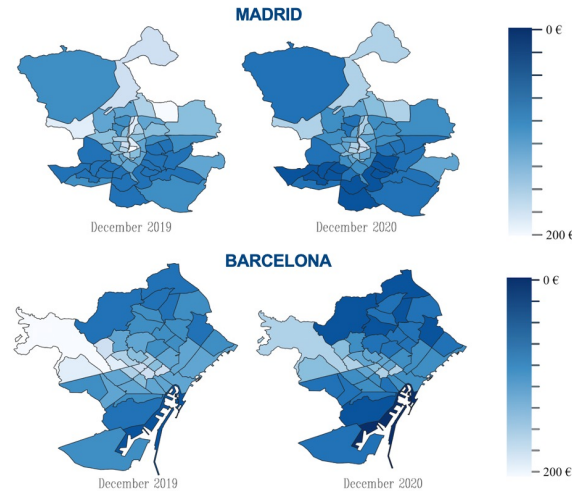
Zooming Out: National, Regional, Provincial, Urban and Zip Code...

Urban Big Data Consumption: Barcelona & Madrid in Real Time & High Definition for Smart Policies

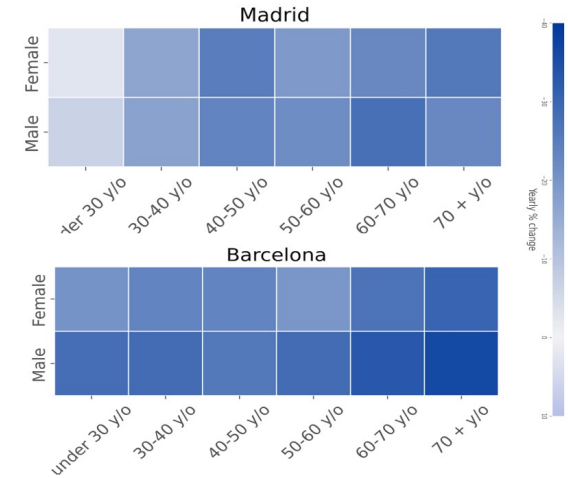
Urban Big Data Consumption: Barcelona & Madrid
(Moving Average 28D. YoY Nominal)



Consumption in Restaurants & Hotels in 2020
(Avg Consumption per person December 2020 & December 2019, %)



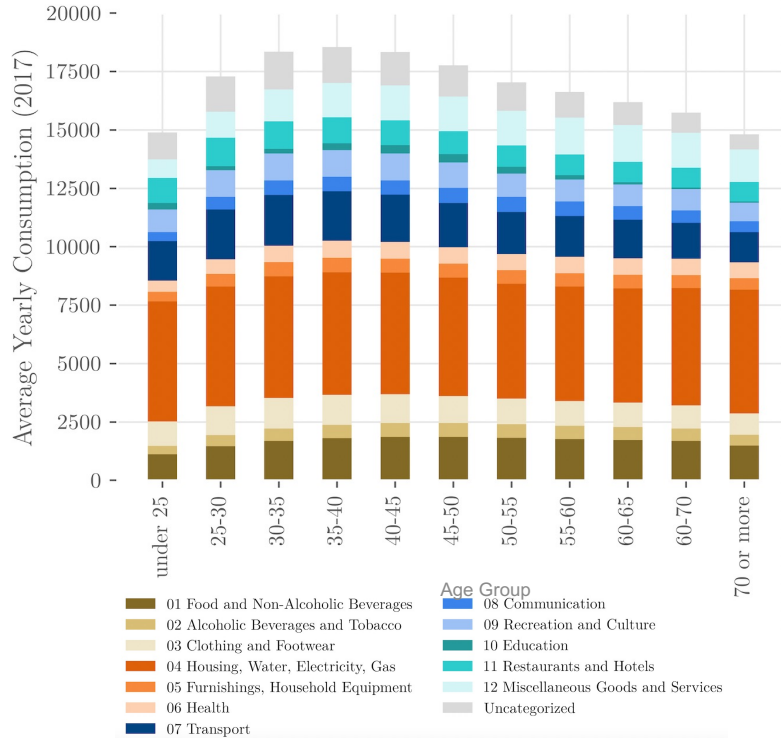
Change in Consumption Restaurants & Hotels in 2020
(December 2020 vs December 2019, % YoY)



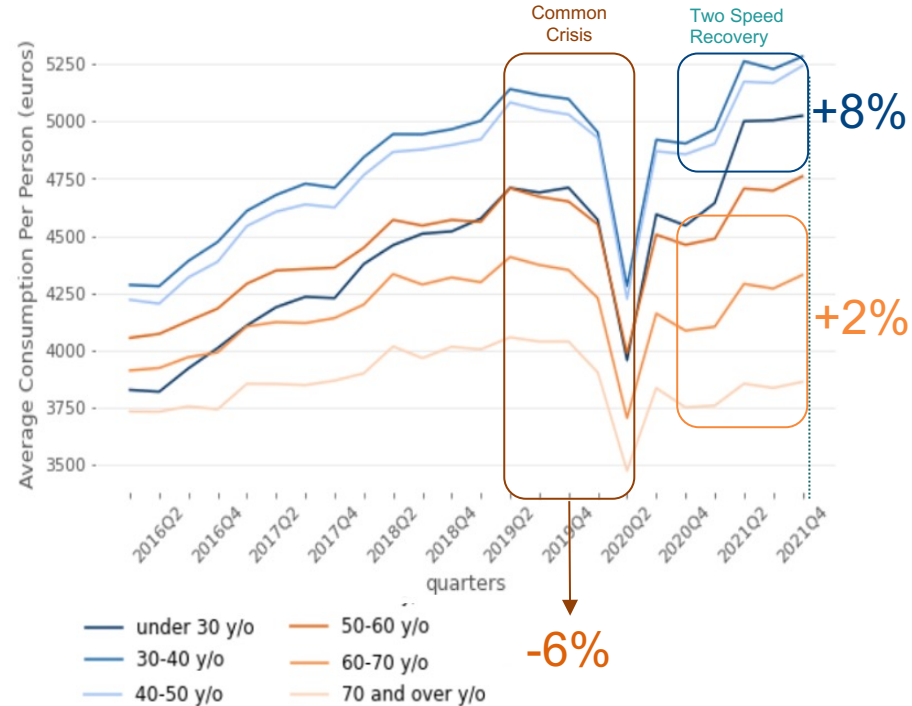
Source: Buda et al (2022) BBVA Research

...Naturally Occurring Data are also relevant for the design of Smart Policies

Spain: Proportion of Consumption per Age Groups (2017)

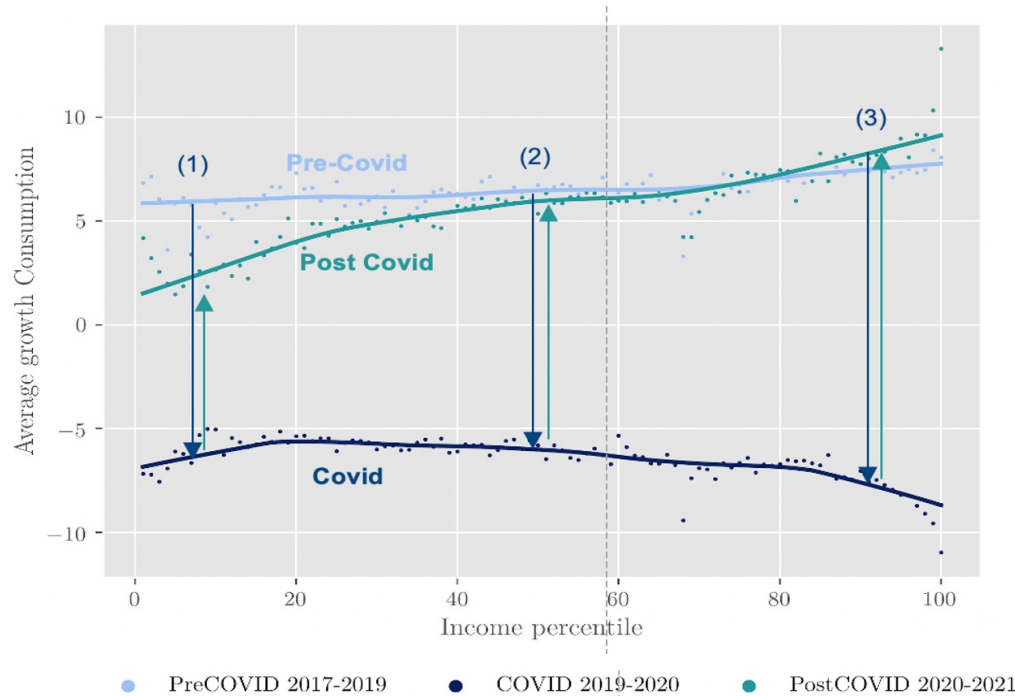


Spain: Levels of Consumption by Age Groups 2016-21



... as High Granularity can give us a “detailed” vision of reality

Spain: Consumption growth by Income Percentile



...Just a partial recovery for the Lower income people

...A sharp adjustment in Consumption Growth for “High Income” and the lower ones ...

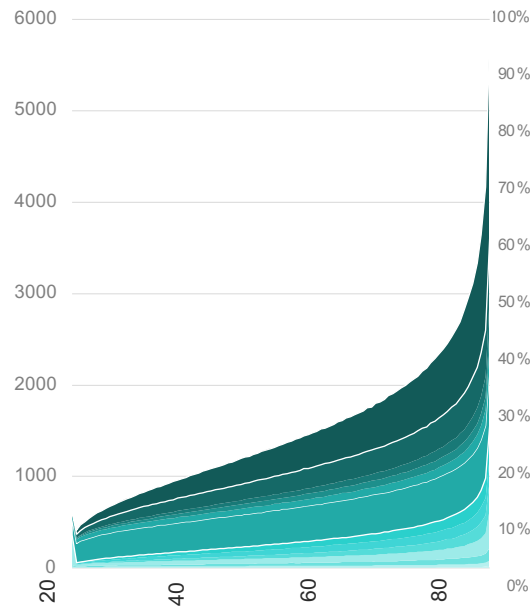
...Remember they are growth rates.. so Consumption Level recovery Incomplete at the end of 2021

and Distributional Consumption insights for Sustainability Policies

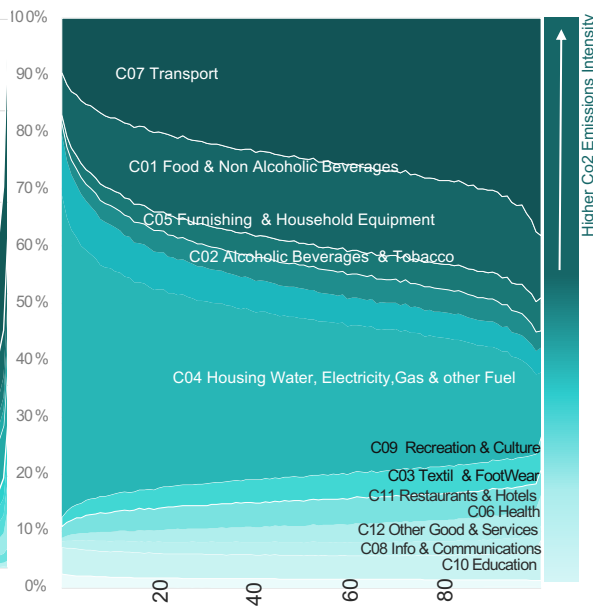
Spain: Co2 Intensity by Category & Percentile & Age groups (2017)

(Percentile of Consumption by CO2 Intensity in EUR, % of total consumption and Age)

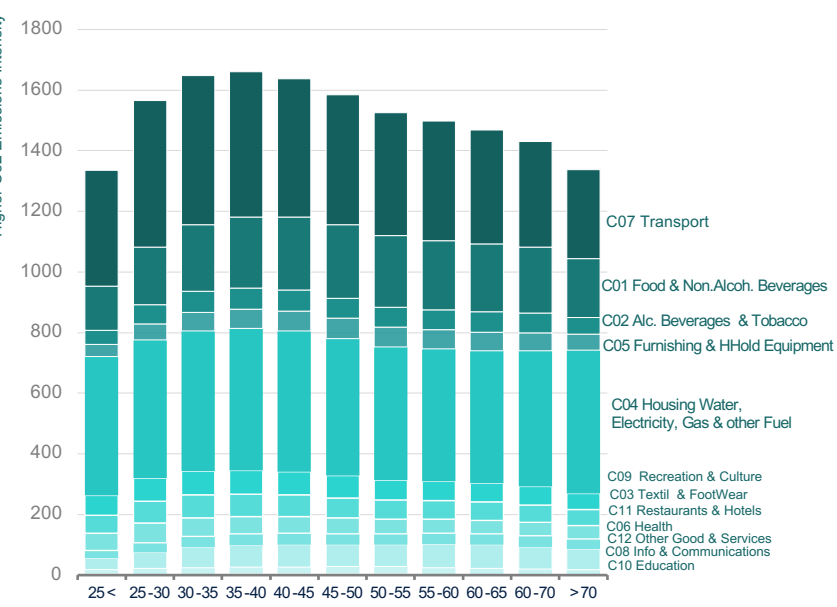
COICOP Category by percentil in EUR



COICOP Category by percentil in % of Total



COICOP Category by percentil and Age in EUR



Source: Buda et al (2022) and BBVA Research. Applying CO2 Intensities to COICOP categories

Conclusions

Conclusions

- This paper provides **the first proof of concept that naturally occurring transaction data, can be harnessed to produce high quality consumption survey.**
- **A simple aggregation of the data can result in good proxies of National Accounts levels and growth**
- The real time and high-definition components of the data show a **powerful economic tool to track the economy in real time not only at aggregate levels but also zooming out across characteristics, gender, income categories and geographical details.**
- The combination of this information constitutes **the basis for better diagnoses and furthermore for the design of smart policies.**

We aim to extend this analysis to other countries and design similar procedures to determine Investment, external sector, internal flows, I-O tables...

National Accounts in a World of Naturally Occurring Data: A Proof of Concept for Consumption

Gergely Buda (BSE)

Vasco M. Carvalho (University of Cambridge & CEPR)

Stephen Hansen (Imperial College & CEPR)

Alvaro Ortiz (BBVA Research)

Tomasa Rodrigo (BBVA Research)

José V. Rodríguez Mora (University of Edinburgh & CEPR)

Conference on Non-traditional Data, Machine Learning and NLP in Macroeconomics, Co-Hosted by
Sveriges Riksbank, Federal Reserve Board, the Bank of Italy and the Bank of Canada.

Stockholm, October 3-4, 2022

Link to [CEPR paper](#)

Link to [Cambridge WP in Economics](#)